

损坏容忍的数据查询降级服务机制

李玲¹ 秦小麟¹ 戴华²

(南京航空航天大学计算机科学与技术学院 南京 210016)¹ (南京邮电大学计算机学院 南京 210003)²

摘要 传统的数据库安全机制重点关注数据的机密性,忽略了用户对数据库系统提供数据的完整性和可用性要求。为了提高系统数据的可用性和用户查询服务的满意度,提出了一种损坏容忍的数据查询降级服务机制。首先,在现有可生存性研究的基础上,引入降级服务的数据模型、数据完整度等相关概念,给出查询降级服务模型的定义;其次,根据该降级服务模型的定义给出相关的查询处理机制与具体查询算法,并从理论上证明该机制的正确性;最后,通过实验从查询执行性能的角度进一步说明该查询降级服务机制的有效性。

关键词 数据库安全,数据库可生存性,数据库隔离,查询降级服务规则

中图分类号 TP392 **文献标识码** A

Damage-tolerant Date Query Degraded Service Mechanism

LI Ling¹ QIN Xiao-lin¹ DAI Hua²

(College of Information Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)¹

(College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)²

Abstract Traditional database security mechanism which focuses on the data confidentiality often overlooks the requirement of data integrity and availability. This paper presented a Damage-Tolerant Data Query Degraded Service (DT-DQDS), for improving system availability and user satisfaction of query operation. First, the model of the degraded service was given, which includes the definitions of data model and data integrity concept based on the existing study of survivability. Second, according to DT-DQDS model, data query mechanism and a concrete query algorithm were provided, and the mechanism is validated theoretically. Finally, experimental results further demonstrate its good performance on the aspect of query execution efficiency.

Keywords Database security, Database survivability, Database isolation, Degraded query service rule

1 引言

数据库管理系统(Database Management System, DBMS)承载着存储和管理关键数据、处理核心计算的重要任务,因此,其安全性格外重要。数据库安全机制重点关注数据库数据的机密性(Confidentiality)、完整性(Integrity)和可用性(Availability)^[1]。传统 DBMS 提供了大量的机制(如认证、权限管理、访问控制、加密技术等^[2,3])来保证数据的机密性,但是这些以预防为主的安全机制无法适应如今多变的攻击环境,无法满足关键 DBMS 系统基本服务不可中断的可生存性^[4]要求。当系统中的数据遭受入侵而受损时,如果能够为用户提供一种“降级”的数据访问服务,一方面能保证查询请求的完整性,提高系统数据的可用性;另一方面也能保证数据库基本服务的持续性。系统提供这种“降级”的服务需要有基本的前提条件——损坏数据隔离(Damage Quarantine)^[5-8],以保证任何用户都无法访问到被污染的数据,导致损坏传播(Damage Spreading)^[9]。

损坏数据隔离的主要思想是在损坏数据被修复之前,阻止其他用户的一切读、写的访问请求,虽然保证了用户获取数据的正确性,但同时也导致了数据库系统数据的低可用性。在文献[8]中 Liu 等提出了一种基于事务依赖的多阶段隔离策略,由于在起始阶段就隔离所有被修改的数据,同时又限制合法用户对于隔离数据的访问,这使得合法且被误隔离的数据也无法访问,大大降低了数据库系统中数据的可用性。文献[10-13]等针对这种过度隔离策略进行了改进,实现了精确隔离,但现有的这些隔离技术往往对于数据的访问策略限制都非常严格,即使达到了精确隔离,数据库系统也无法容忍包含损坏数据的访问服务,而多采取直接拒绝服务请求的策略。这种策略破坏了用户请求数据的完整性,无法保证系统数据的高可用性,并且带来糟糕的用户查询服务体验。

针对这些问题,本文提出了一种损坏容忍的数据查询降级服务机制,解决了数据隔离技术带来的可用性低的问题。该机制允许包含损坏数据的查询操作,返回用户“降级”的查询结果集合,而非直接拒绝或丢弃合法用户的查询请求。这

到稿日期:2012-08-30 返修日期:2012-12-03 本文受国家自然科学基金项目(60673127),国家 863 高技术计划项目(2007AA01Z404),博士点基金项目(20103218110017)资助。

李玲(1987-),女,硕士生,主要研究方向为数据库安全、数据库可生存性技术, E-mail: tracylling@gmail.com; 秦小麟(1953-),教授,博士生导师,主要研究方向为时空数据库、分布式环境的数据管理与安全、信息安全;戴华(1982-),博士,讲师,主要研究方向为传感器网络数据管理与安全、数据库安全。

种服务机制保证了数据库系统的服务功能不中断,且为用户提供了满足期望的未受损数据对象集合,并带来系统服务可用性的提高以及数据完整性的提高。本文首先给出损坏容忍的数据查询降级服务的相关模型定义和服务规则定义,提出数据查询降级服务的算法;然后定义查询执行成功率,以对该服务模型的有效性进行分析;最后给出该模型在查询执行性能方面的实验分析。

2 损坏容忍的查询降级服务

我们假设此时入侵行为已经发生并被检测,被损坏及被污染的数据已被有效隔离。本文中所有被恶意事务影响的事务均为污染事务,被污染事务所修改的数据均为受损数据。

2.1 查询降级服务的数据模型

定义 1 一个数据库是一组数据对象的集合,表示 $DB = \{x_1, x_2, \dots, x_n\}$ 。为了描述的简洁性,这里的数据对象都简称为“数据”。

定义 2 对于数据库中的任意数据 x ,存在两个属性 ($val, flag$):

(1) val 表示数据 x 的值,记为 $x.val$;

(2) $flag$ 为数据 x 是否为损坏数据对象标记,记为 $x.flag, flag \in \{TRUE, FALSE\}$,若 $x.flag = TRUE$,则 x 为损坏数据;若 $x.flag = FALSE$,则 x 为合法数据。

定义 3 对于用户查询请求 q ,将 q 的查询结果数据集记为 $\Omega(q)$,根据数据是否损坏,可以将 $\Omega(q)$ 分为如下两个部分:

(1) 合法结果集 (Valid Result Set),即查询结果中合法数据构成的集合,记为 $\mathcal{R}(q)$,则

$$\mathcal{R}(q) = \{x | x \in \Omega(q) \wedge x.flag = FALSE\}$$

(2) 非法结果集 (Invalid Result Set),即查询结果中损坏数据构成的集合,记为 $\mathcal{I}(q)$,则

$$\mathcal{I}(q) = \{x | x \in \Omega(q) \wedge x.flag = TRUE\}$$

我们用“ $|S|$ ”表示数据集 S 中元素的个数,称其为集合的秩。

对于用户的查询请求,只有合法结果集中的数据才会返回给用户,用户访问不到非法结果集中的数据,这样保证了数据的有效性,同时也有效地抑制了受损数据的传播。

定义 4 数据可用性是指数据库中数据的可用程度,本文使用数据查询的完整度来进行描述,即对于用户的查询请求,数据库能够返回查询结果的完整性。主要可分为以下两类完整度定义:

(1) 期望完整度 (Expect Integrity, EI),它表示用户请求对于数据库返回查询结果的完整度的期望值,记为 α 。

(2) 实际完整度 (Practical Integrity, PI),它表示实际数据库系统对于用户请求所返回的实际查询结果完整度,假设用户提交的查询记为 q ,则实际完整度可用 $\delta(q)$ 表示,有计算公式如下

$$\delta(q) = \frac{|\mathcal{R}(q)|}{|\mathcal{R}(q)| + |\mathcal{I}(q)|} = \frac{|\mathcal{R}(q)|}{|\Omega(q)|}$$

定义 5 损坏容忍的数据查询降级服务 (Damage-Tolerant Data Query Degraded Service, 简称 DT-DQDS) 是这样一种服务策略:用户可以提交包含损坏数据的查询请求而不被系统直接拒绝,系统判断是否返回给用户“降级”的查询结果。判断过程遵循服务描述中定义的相关规则,最终形成动态的

服务交互过程,并返回满足用户期望完整度的数据结果集合。

2.2 数据查询降级服务规则

基于损坏容忍的数据查询降级服务的基本思想,我们给出服务中的相关规则描述。

定义 6 数据查询降级服务规则 (Degraded Query Service Rule, DQSR):对于用户查询 q , q 的执行与否应满足查询降级服务规则,该规则包含执行查询规则 (Executing Query Rule, EQR) 和拒绝查询规则 (Denying Query Rule, DQR):

(1) EQR 规则:如果查询 q 满足条件 $\delta(q) \geq \alpha$,则系统执行查询 q ,并返回用户合法查询结果集 $R(q)$ 和相应的实际完整度值 $\delta(q)$ 。

(2) DQR 规则:如果查询 q 满足条件 $\delta(q) < \alpha$,则系统拒绝查询 q 的执行,并将其挂起,等待修复程序的数据修复过程。

由此,与传统的可生存性技术中的查询服务策略(即当用户查询请求访问到一条损坏数据时,直接拒绝用户服务)相比,该机制只需满足 EQR 规则,便可返回用户部分未受损的合法数据,提高了查询操作执行的成功率,也提高了系统可用性和用户请求数据的完整性。

2.3 降级服务算法实现

损坏容忍的数据查询降级服务算法 (Damage-Tolerant Data Query degradation Service Algorithm, 简称 DT-DQDSA) 通过比较查询结果的实际完整度 PI 与期望完整度阈值 α ,来确认是否需要将部分满足查询请求的合法数据返回给用户。为方便描述,我们定义查询结果的结构体 R 为二元组 (ds, ir),其中 ds 表示返回给查询用户的结果数据集,记为 $R.ds$; ir 表示返回结果的实际完整度,记为 $R.ir$ 。具体 DT-DQDSA 算法的过程如算法 1 所述。

算法 1 DT-DQDSA 算法

输入:用户查询请求 q

输出:查询结果或退出查询

算法流程:

1. 初始化合法数据计数器 $VQ = 0$,非法数据计数器 $IQ = 0$;
2. 初始化查询结果结构体 R ,使得 $R.ds = \{\}$, $R.ir = 0$;
3. FOR EACH x 满足 q 的查询条件 DO
4. IF $x.flag = FALSE$ THEN
5. $VQ++$; //合法数据个数加 1
6. $R.ds = R.ds + \{x\}$;
7. ELSE IF $x.flag = TRUE$ THEN
8. $IQ++$; //非法数据个数加 1
9. END IF
10. END DO
11. $R.ir = VQ / (VQ + IQ)$ //计算查询结果的实际完整度
12. IF $R.ir \geq \alpha$ THEN
13. RETURN R ;
14. END IF
15. RETURN NIL; //不满足置信度阈值,查询失败

在算法 1 中,首先扫描数据库中的每个数据项,在满足查询条件的情况下,根据损坏数据标志位将结果分为合法和非法数据集,如步骤 4,7 所示;同时将值返回至两类集合中,计算集合的大小。程序扫描完最后一个数据项后,计算实际完整度 ir 并与期望完整度 α 进行比较。若满足 ERQ 规则,则返回用户合法数据集,否则拒绝用户的请求,查询失败。

3 查询执行成功率评估方法

我们提出查询执行成功率的定义,将其作为 DT-DQDS

机制对数据库查询性能的评价指标。

定义 7 查询执行成功率 (Query Execution Success Rate) 是指数据库查询操作能够成功执行并满足用户期望返回结果集的概率, 若将查询请求记为 Q , 则查询请求 Q 的查询执行成功率记为 $P(Q)$ 。

我们不妨假设 $|DB|$ 表示数据库中数据对象的总数量。下面给出相关符号定义, 如表 1 所列。

表 1 相关符号说明

符号	说明
N	请求 Q 访问数据对象的数量
\mathcal{R}	查询结果中合法数据集
\mathcal{S}	查询结果中非法数据集
α	期望完整度 EI
p	损坏数据百分比
X	数据库中合法数据量, 即 $ DB * (1-p)$
Y	数据库中非法数据量, 即 $ DB * p$

在现有的数据查询机制中, 系统绝对禁止任何用户提交包含损坏数据的查询请求, 采用拒绝服务的方式。可知, 用户执行查询操作获得的数据必定都为合法数据集中的数据项, 查询所访问的 N 个数据项都为合法数据。由定义 4 可知, 用户对于数据库返回查询结果的期望值为 1, 意味着, 当 α 取值为 1 时, 则传统可生存性技术研究中的数据查询的执行成功率如式(1)所示。

$$P_T(Q) = \frac{C_X^N C_Y^0}{C_{|DB|}^N}, N \leq X \quad (1)$$

在现有研究技术的基础上, 我们接着定义 DT-DQDS 模型中的查询执行成功率。由上述定义 5 中的 DQSR 规则可得以下式子成立:

$$|\mathcal{R}| + |\mathcal{S}| = N \quad (2)$$

$$\delta(Q) = \frac{|\mathcal{R}|}{N} \geq \alpha \quad (3)$$

在满足上述两式的情况下, 我们定义 DT-DQDS 模型中查询请求 Q 的执行成功率 $P(Q)$, 可分为以下两种情况:

(1) 当 $N \leq X$, 即访问的数据对象数量 N 小于整个数据库 DB 中所有合法数据对象数量 X 时, 可以得到查询执行成功率如下:

$$P(Q) = \frac{C_X^{[aN]} C_Y^{N-[aN]}}{C_{|DB|}^N} + \frac{C_X^{[aN]+1} C_Y^{N-[aN]-1}}{C_{|DB|}^N} + \dots + \frac{C_X^N C_Y^0}{C_{|DB|}^N} \\ = \sum_{i=[aN]}^N \frac{C_X^i C_Y^{N-i}}{C_{|DB|}^N} \quad (4)$$

(2) 当 $N > X$, 即访问的数据对象数量 N 超过整个数据库 DB 中所有合法数据对象数量 X , 所访问的数据对象中至少有一项为非法数据集 Y 中的数据时, 可以得到查询执行成功率, 表示如下:

$$P(Q) = \frac{C_X^{[aN]} C_Y^{N-[aN]}}{C_{|DB|}^N} + \frac{C_X^{[aN]+1} C_Y^{N-[aN]-1}}{C_{|DB|}^N} + \dots + \frac{C_X^X C_Y^{N-X}}{C_{|DB|}^N} \\ = \sum_{i=[aN]}^X \frac{C_X^i C_Y^{N-i}}{C_{|DB|}^N} \quad (5)$$

由于期望完整度 $\alpha \in [0, 1]$, 因此查询成功率中合法数据集 \mathcal{R} 的大小取值范围为 $[N, N \times \alpha]$, 依次列举该范围中所有可能发生的情况, 进行概率累加, 即可获得在确定期望完整度阈值下的本次查询操作的执行成功率。由式(4)、式(5)构成 DT-DQDS 模型最终的查询执行成功率表达式。

在 DT-DQDS 模型中, 我们加入了可变因素期望完整度 α , 对最终的查询执行成功率产生了比较重要的影响。同时,

数据库的总规模、数据库中损坏率的大小和用户查询访问数据量的大小对查询性能都会产生一定的影响, 具体由下文实验进行分析和评估。

4 实验结果分析

为了进一步说明 DT-DQDS 模型的有效性, 将从查询执行成功率的角度, 对传统的数据查询机制和 DT-DQDS 模型中的数据查询机制进行查询执行性能的对比实验。实验采用的硬件环境为 Intel Core(TM) i5 处理器, 4G 内存; 软件环境为 Ubuntu10.04 操作系统、MySQL 后台数据库和 Python 脚本开发工具。

我们选取数据库的总数据量(元组数量)为 5000, 在对参数 α 、 N 以及 X 进行不同设置的情况下, 分析各参数对于查询执行成功率的影响, 具体实验分为以下 3 组:

(1) 当 $|DB| = 5000$ 时, 以期望完整度 α 为自变量 ($0 < \alpha \leq 1$), 实验结果如图 1 所示。

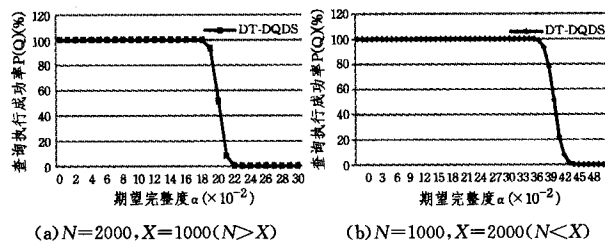


图 1 期望完整度 α 为自变量时的查询执行成功率 $P(Q)$

由图 1 可知, 在 DT-DQDS 服务机制下, 当期完整度 α 在某范围内时, 系统获得趋近 100% 的查询执行成功率, 即提供完全满足用户需求的降级的查询服务; 随着 α 的增加, $P(Q)$ 最后迅速趋于 0%, 系统将无法再提供查询服务。该范围取决于数据库中合法数据的比例, 即 X 与 $|DB|$ 的比值, 该比值又可表示为从系统中取得一条数据为合法数据的概率。

由式(1)可知, 在本组实验参数 N 与 X 的取值下, 传统查询机制仅可获得趋于 0% 的查询执行成功率。实验表明, DT-DQDS 服务机制提升了系统整体的查询执行成功率, 能够更好地提高用户服务满意度和系统数据的可用性, 且数据库系统的数据损坏率越小, DT-DQDS 服务提升的执行性能越高。

(2) 当 $|DB| = 5000$ 、 $X = 3000$ 时, 以查询访问量 N 为自变量 ($500 \leq N \leq 4500$), 随着 N 的变化, 查询执行成功率 $P(Q)$ 的变化如图 2 所示。

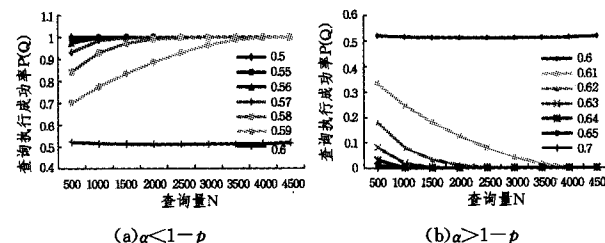


图 2 查询访问量 N 在不同期望完整度 α 下对于查询执行成功率 $P(Q)$ 的影响

由式(1)可知, 传统查询机制的查询执行成功率与 α 无关, 随着 N 的增加, 其查询执行成功率均无限趋近于 0%。而在 DT-DQDS 服务机制下, 期望完整度 α 在不同取值范围下, 查询访问量 N 作为自变量时对查询执行成功率的影响有明显区别。由图 2(a) 可知, 当期完整度 α 小于合法数据量比

例,即 $\alpha < 1-p$ 时,随着 N 的增加, $P(Q)$ 逐渐提高,趋向于 100%, α 取值越小, $P(Q)$ 越快地趋于 100%; 相反地,在图 2 (b) 中,当 $\alpha > 1-p$ 时,随着 N 的增加, $P(Q)$ 逐渐降低,并逐渐趋向于 0%,且 α 取值越大, $P(Q)$ 越快地趋于 0; 当 $\alpha = 1-p$ 时,随着查询量 N 的增加,查询执行成功率 $P(Q)$ 保持在 0.5 以上,呈一条光滑的曲线。

该组实验结果表明,用户期望度越低,查询请求被拒绝的可能性就越低,所以能够获得较高的查询执行成功率;相反,期望度越高,查询操作越容易拒绝;期望度越接近合法数据所占比例,查询执行成功率的变化趋势越平缓。

(3) 当 $|DB| = 5000, N = 2000$ 时,以 X 为自变量 ($500 \leq X < 5000$),随着 X 的变化,查询执行成功率 $P(Q)$ 的变化如图 3 所示。

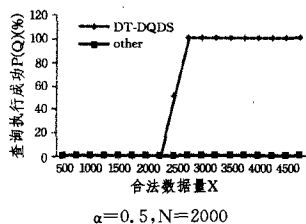


图 3 合法数据量 X 为自变量时的查询执行成功率 $P(Q)$

由图 3 可知,当合法数据量 X 逐渐增大时,查询执行成功率由 0% 上升至 100%。这是由于在 DT-DQDS 服务机制下,当用户期望完整度确定时,若数据库系统中合法数据量越大,数据损坏率就越小,所以查询到合法数据的几率增大,综合导致最后的查询执行成功率增长。

综合上述 3 组实验可知,在实验设置的可变因素下,DT-DQDS 查询降级服务机制与传统的可生存性环境中的查询服务机制相比,具有更好的查询性能。在一定用户定义期望完整度阈值下,DT-DQDS 降级服务仍能够为用户提供数据查询服务,在保证数据准确性的基础上,数据请求的完整性也得到了提高,相比传统查询机制 0% 的查询执行成功率,DT-DQDS 能维持一定比例的近 100% 的查询执行成功率,很大程度上提高了数据的可用度和用户的满意度。

结束语 为了解决现有可生存性数据库中查询机制带来的数据可用性低的问题,本文提出了一种损坏容忍的降级服务模型 DT-DQDS。首先给出了 DT-DQDS 的定义和服务规则 DQSR,并根据服务规则 DQSR,给出了基于该模型的查询服务机制的核心算法。其思想是通过 DQSR 规则进行判断,以决定是否返回用户部分合法的数据集合,即降级的查询结果。该机制避免了系统对于包含损坏数据查询访问的完全拒绝,从而实现了系统数据可用性的提高,提高了用户的查询服务体验。实验结果表明,DT-DQDS 服务机制相比现有的数

据查询服务机制,当用户设置的期望度阈值小于系统合法数据的比例时,DT-DQDS 机制下的查询操作成功率将大幅提高,获得趋近 100% 的执行成功率(而传统查询机制执行成功率仅为 0%),其系统可用性也得到显著提高。所以 DT-DQDS 机制在入侵发生后能够提供更好的数据查询服务。

参考文献

- [1] Bertino E, Sandhu R. Database security-concepts, approaches, and challenges [J]. IEEE Transactions on Dependable and Secure Computing, 2005, 2(1): 2-19
- [2] 刘启原, 刘怡. 数据库与信息系统的的天 [M]. 北京: 科学出版社, 2000
- [3] 张敏, 徐震, 冯登国. 数据库安全 [M]. 北京: 科学出版社, 2005
- [4] Knight J C, Sullivan K J, Elder M C, et al. Survivability architectures: Issues and approaches [C]// DARPA Information Survivability Conference and Exposition, 2000 DISCEX '00 Proceedings. 2000: 157-171
- [5] Ammann P, Jajodia S, McCollum C D, et al. Surviving information warfare attacks on databases [C]// Security and Privacy, 1997 Proceedings of IEEE Symposium on. 1997: 164-174
- [6] Bai K, Liu P. A data damage tracking quarantine and recovery (DTQR) scheme for mission-critical database systems [C]// Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology. ACM, Saint Petersburg, Russia, 2009: 720-731
- [7] Bai K, Yu M, Liu P. Trace: Zero-down-time database damage tracking, quarantine, and cleansing with negligible run-time overhead [C]// Proceedings of the 13th European Symposium on Research in Computer Security. Computer Security. Springer-Verlag, Málaga, Spain, 2008: 161-176
- [8] Liu P, Jajodia S. Multi-phase damage confinement in database systems for intrusion tolerance [C]// Computer Security Foundations Workshop, 2001. Proceedings. 14th IEEE, 2001: 191-205
- [9] Bai K, Liu P. Towards database firewall: Mining the damage spreading patterns [C]// Computer Security Applications Conference, ACSAC '06 22nd Annual. 2006: 449-462
- [10] 戴华, 秦小麟, 郑吉平. 基于 CTMO 模型的数据库损坏数据隔离技术 [J]. 计算机学报, 2011, 34(2): 275-290
- [11] 谢美意, 朱虹, 冯玉才, 等. 自修复数据库系统设计实现关键问题研究 [J]. 小型微型计算机系统, 2010(10): 1926-1930
- [12] 付戈, 时杰, 李专, 等. 一种有效的受损数据隔离方法 [J]. 计算机科学与探索, 2010(8): 712-722
- [13] Dai H, Qin X, Zheng G, et al. SQRM: An Effective Solution to Suspicious Users in Database [C]// Proceedings of 3rd International Conference on Advances in Databases, Knowledge, and Data Applications. 2011: 59-64
- [14] density [J]. Probabilistic Networks and Expert Systems, 1999, 172(35): 163-178
- [15] Er M J. Genetic algorithms for MLP neural network parameters optimization [C]// Proceedings of the Chinese Control and Decision Conference, 2009. Guilin, IEEE, 2009: 3653-3658
- [16] Burges C. A tutorial on SVM for pattern recognition [J]. Data Mining and Knowledge Discovery, 1998, 10(2): 121-167
- [17] Plutchik R, Kellerman H. A general psychoevolutionary theory of emotion [J]. Theories of Emotion, 1980, 1(3): 3-33
- [11] Lalande A, Legrand L, Walker P M, et al. Automatic detection of cardiac contours on MR images using fuzzy logic and dynamic programming [C]// Proceedings of the AMIA Annual Fall Symposium, 1997. Nashville, AMIA, 1997: 474-478
- [12] Zahlmann G, Scherf M, Wegner A. A neuro fuzzy classifier for a knowledge-based glaucoma monitor [C]// Lecture Notes in Artificial Intelligence, 1997. Germany, Springer, 1997: 273-287
- [13] Cowell R, Dawid A, Lauritzen S, et al. Psychologic pharmacokinetics model based on Bayes network with optimal of kernel

(上接第 66 页)